

This is a repository copy of *Long-term reciprocal gene flow in wild and domestic geese reveals complex domestication history*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/163663/>

Version: Accepted Version

Article:

Heikkinen, Marja, Ruokonen, Minna, White, Thomas et al. (6 more authors) (2020) Long-term reciprocal gene flow in wild and domestic geese reveals complex domestication history. G3: Genes, Genomes, Genetics. ISSN 2160-1836

<https://doi.org/10.1534/g3.120.400886>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Long-term reciprocal gene flow in wild and domestic geese reveals complex domestication history

Marja E. Heikkinen,^{*,†,1} Minna Ruokonen^{*,2} Thomas A. White,^{*,‡} Michelle M. Alexander,[§] İslam Gündüz,^{**} Keith M. Dobney,^{††,‡‡,§§} Jouni Aspi,^{*} Jeremy B. Searle,^{†,3} Tanja Pyhäjärvi^{*,1,3}

^{*}Department of Ecology and Genetics, PO Box 3000, FI-90014 University of Oulu, Finland

[†]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA

[‡]CMPG, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland

[§]University of York, BioArCh, Environment Building, Wentworth Way, Heslington, York, YO10 5NG, UK

^{**}Department of Biology, Faculty of Arts and Sciences, University of Ondokuz Mayıs, Samsun, Turkey

^{††}Department of Archaeology, Classics and Egyptology, University of Liverpool, 12–14 Abercromby Square, Liverpool L69 7WZ, UK

^{‡‡}Department of Archaeology, University of Aberdeen, St Mary's, Elphinstone Road, Aberdeen, AB24 3UF, UK

^{§§}Department of Archaeology, Simon Fraser University, Burnaby, B.C. V5A 1S6, 778-782-419, Canada

22 Short running title: Gene flow in wild and domestic geese

23 Keywords: *Anser anser*, domestication, domestic goose, hybridization, population genomics

24 ¹ Co-corresponding authors:

25 Marja E. Heikkinen

26 Department of Ecology and Genetics

27 PO Box 3000

28 Fi-90014 University of Oulu

29 Finland

30 Phone number +358408275850

31 marja.e.heikkinen@oulu.fi

32

33 Tanja Pyhäjärvi

34 Department of Ecology and Genetics

35 PO Box 3000

36 Fi-90014 University of Oulu

37 Finland

38 Phone number: +358503504932

39 tanja.pyhajarvi@oulu.fi

40

41 ² Deceased.

42 ³ Contributed equally to this work.

43

44

45 **Abstract**

46 Hybridization has frequently been observed between wild and domestic species and can substantially
47 impact genetic diversity of both counterparts. Geese show some of the highest levels of interspecific
48 hybridization across all bird orders, and two of the goose species in the genus *Anser* have been
49 domesticated providing an excellent opportunity for a joint study of domestication and hybridization.
50 Until now, knowledge of the details of the goose domestication process has come from archaeological
51 findings and historical writings supplemented with a few studies based on mitochondrial DNA. Here,
52 we used genome-wide markers to make the first genome-based inference of the timing of European
53 goose domestication. We also analyzed the impact of hybridization on the genome-wide genetic
54 variation in current populations of the European domestic goose and its wild progenitor: the graylag
55 goose (*Anser anser*). Our dataset consisted of 58 wild graylags sampled around Eurasia and 75
56 domestic geese representing 14 breeds genotyped for 33,527 single nucleotide polymorphisms.
57 Demographic reconstruction and clustering analysis suggested that divergence between wild and
58 domestic geese around 5,300 generations ago was followed by long-term genetic exchange, and that
59 graylag populations have 3.2–58.0% admixture proportions with domestic geese, with distinct
60 geographic patterns. Surprisingly, many modern European breeds share considerable (> 10%)
61 ancestry with the Chinese domestic geese that is derived from the swan goose *Anser cygnoid*. We
62 show that the domestication process can progress despite continued and pervasive gene flow from
63 the wild form.

64 1. Introduction

65 Reproductive isolation is a defining feature of speciation and yet hybridization between species is an
 66 important general phenomenon in evolution (Arnold 2004; Abbott *et al.* 2013). Among birds, the
 67 Anseriformes (ducks, geese, and swans) show particularly pervasive hybridization, 41.6% to > 60% of
 68 species hybridizing with each other (Grant and Grant 1992; Ottenburghs *et al.* 2016a). Domestication
 69 generates differentiated gene pools and reproductive isolation between domestics and their wild
 70 progenitor, but hybridization between domestic and wild forms has been well demonstrated in both
 71 plants (Arnold 2004; Janzen *et al.* 2019) and animals (Godinho *et al.* 2011; Frantz *et al.* 2015). The
 72 impacts include genetic and trait enrichment of domestics, for instance, in chicken the acquisition of a
 73 yellow skin phenotype is a result of past mating between red junglefowl and grey junglefowl (Eriksson
 74 *et al.* 2008). In geese, a high tendency for hybridization between wild and domestic forms has also
 75 been suggested (Kuijken and Devos 1996; Heikkinen *et al.* 2015), creating an exciting opportunity to
 76 study the complex dynamics of hybridization and domestication.

77 The domestic geese of the world (European and Chinese forms) are derived from two different
 78 wild species: the graylag (*Anser anser*) and the swan goose (*Anser cygnoid*), respectively (Delacour
 79 1954; Shi *et al.* 2006). *A. anser* and *A. cygnoid* shared a common ancestor about 3.4 Mya
 80 (Ottenburghs *et al.* 2016b) but are still able to hybridize (Ottenburghs *et al.* 2016a), and some
 81 domestic breeds are reportedly hybrid (Buckland and Guy 2002). The graylag has been divided into
 82 the western, nominate subspecies *A. a. anser* (Linnaeus, 1758) with a European breeding range and
 83 the eastern subspecies *A. a. rubrirostris* (Swinhoe, 1871) breeding further east, although the
 84 subspecific boundary is not well defined, and mitochondrial DNA has not been found to distinguish
 85 them (Heikkinen *et al.* 2015). Of these subspecies, *rubrirostris* is larger and lighter colored than *anser*
 86 (Cramp and Simmons 1977) and has a pink bill and cold pink legs in contrast to the orange bill and

87 flesh-colored legs of *anser*, the bill color used as primary evidence in favor of the original
88 domestication of *rubrirostris* (Kear 1990). As with all domesticates, domestic geese varieties are
89 morphologically more diverse than their wild counterparts, particularly in plumage and body size
90 (Buckland and Guy 2002).

91 The current knowledge about goose domestication relies largely on ancient texts and
92 archaeological evidence. Questions about where and when domestication took place, the genetic
93 changes associated with it and the later history of domestic geese, however, remain largely
94 unresolved (Heikkinen *et al.* 2015). There are depictions from the New Kingdom of Egypt that suggest
95 geese were already fully domesticated by the 18th Dynasty (1450-1341 BCE). The earliest reliable
96 reference to domestic geese in western Eurasia is Homer's *Odyssey* (first half of 8th century BCE) and
97 geese were certainly well-established poultry by Roman times (Albarella 2005).

98 Genetic diversity in the mitochondrial DNA (mtDNA) of graylag and European domestic geese
99 showed reduced diversity in the domestics (Heikkinen *et al.* 2015) which may result from an early
100 domestication bottleneck or, alternatively, later breed formation. There is a particular mitochondrial
101 haplogroup common in the domestics (Heikkinen *et al.* 2015), and archaeological domestic goose
102 bones from the High Medieval (11th-13th century CE) of Russia belonged to that haplogroup (Honka *et*
103 *al.* 2018).

104 MtDNA relationships between extant Chinese and European domestic goose breeds confirm
105 that the former, excluding one breed, have swan goose ancestry, whereas European domestic goose
106 and the Chinese Yili breed have graylag ancestry (Shi *et al.* 2006; Sun *et al.* 2013; Ren *et al.* 2016).
107 However, Chinese mtDNA haplotypes may occasionally occur in European domestics, and vice versa
108 (Sun *et al.* 2013; Heikkinen *et al.* 2015).

Genomic data can be much more powerful than mtDNA in terms of inference about hybridization. For instance, New World cattle, along with their taurine ancestry have been shown genomically to have a greater proportion of indicine ancestry than previously assumed (McTavish *et al.* 2013) and genomic studies of domestic pigs have shown them to have received genetic input from wild boars (Frantz *et al.* 2015). Genomic studies of modern dog breeds also show an ancestry that can only be explained by gene flow from multiple regional wolf populations (Skoglund *et al.* 2015). Plant varieties are often shown to be the product of hybridization by genomic studies, for example maize (Hufford *et al.* 2013). Interpretation of genomic data is still challenging and for the study of domestic species and their interactions with their wild progenitors, it is best to apply genomics to infer jointly the genetic impact of initial domestication and subsequent hybridization of wild and domestic populations, as the latter can obscure domestic-wild genetic relationships and may also give a false impression of the location and number of times a species has been domesticated (van Heerwaarden *et al.* 2011; Marshall *et al.* 2014; Larson and Fuller 2014).

Here we investigate goose domestication history using genome-wide single nucleotide polymorphism (SNP) data from thousands of loci, obtained by genotyping-by-sequencing (GBS). We used 56 and 50 samples of graylag and domestic geese from a previous mtDNA study (Heikkinen *et al.* 2015), together with 2 new Turkish graylag and 25 new domestic specimens. We studied the interplay between domestication and hybridization by addressing the following questions: i) what is the extent of genetic differentiation amongst wild and domestic geese? ii) what is the approximate time of domestication? and iii) what is the role of intra- and interspecific hybridization in goose domestication history and iv) how does hybridization affect the genetic composition of modern populations?

2. Materials and methods

a) Sampling

132 The wild-collected graylag samples derive widely from Eurasia (Figure 1, Supplementary File 1, Table
133 S1) representing both subspecies. As no morphological data were available, we could not discriminate
134 the samples between eastern and western subspecies. However, based on their sampling and the
135 known geographic distribution of the populations, we can be confident that the Iranian and
136 Kazakhstani samples belonged to the eastern subspecies *rubrirostris*. The European domestic goose
137 samples represented 14 different breeds (Supplementary File 1, Table S1) together with individuals
138 unattributed to a recognized breed or which were presumptive hybrids between European and
139 Chinese domestic geese. Some specimens were reported to be Chinese domestic geese. The domestic
140 samples were obtained from local breeders in Denmark, Sweden, and the UK, and those from Turkey
141 were collected directly by the authors.

142 b) DNA extraction and GBS library construction

143 GBS (Elshire *et al.* 2011) libraries were constructed at the Cornell Biotechnology Resource Center
144 (BRC) following DNA extraction with the DNeasy Blood and Tissue Kit (QIAGEN) with RNase treatment.
145 Each individual DNA sample and an adaptor with a unique barcode were combined in a 96-well plate
146 along with a common adaptor. Samples were treated with the EcoT-22I (ATGCAT) restriction enzyme
147 to create fragmented DNA. Barcoded adapters and common adapters with matching sticky ends were
148 ligated to each sample with T4 DNA ligase. The samples were pooled and purified with a QIAquick PCR
149 Purification Kit (QIAGEN). PCR amplification of the library used primers complementary to barcoded
150 and common adapters with products purified as above, and the samples were 100 bp SE-sequenced
151 with Illumina HiSeq 2000/2500 at the BRC.

152 c) GBS pipeline and SNP calling

153 Raw sequence reads were run through the Command Line Interface of the Tassel 5 GBS v2 Discovery
 154 and Production pipelines (Glaubitz *et al.* 2014). Details about the pipelines and SNP calling are in the
 155 Supplementary File 1 (see Figure S1 for quick outline of the workflow). Good quality reads were
 156 recorded as tags and aligned to the *A. cygnoid domesticus* GenBank assembly
 157 (AnsCyg_PRJNA183603_v1.0 GCF_000971095.1) (Lu *et al.* 2015) using the Burrows-Wheeler Aligner
 158 with default settings (Li and Durbin 2009). After running the raw data through the pipelines, 69,865
 159 SNPs were obtained.

160 The SNPs were subjected to additional filtering using VCFtools (Danecek *et al.* 2011). We
 161 removed indels, loci with more than two alleles and invariant loci. However, loci that were within-
 162 species invariant but divergent from the reference were retained for phylogenetics, informing about
 163 graylag-swan goose divergence. After preliminary analyses loci with observed heterozygosity over
 164 0.75 were removed as potential paralogs. Individuals with more than 20% missing data across loci
 165 were removed. The final dataset consisted of 33,527 biallelic SNPs and 133 individuals (58 wild and 75
 166 domestic).

167 d) The estimation of genetic diversity

168 Genetic diversity and pairwise F_{ST} values were investigated with the hierfstat R package (Goudet
 169 2005). Expected heterozygosity (H_E) was calculated for each locus and population and averaged across
 170 loci. Difference in average H_E between graylags and European domestics was tested with a two-
 171 sample t-test with the Welch correction for non-homogeneity of variance (Welch 1938). For
 172 comparing the genetic diversity among wild and domestics, only pure graylag populations (defined as
 173 having < 10% admixture with domestic geese) and pure European domestic geese (defined as having <
 174 10% admixture with Chinese domestic geese) were used to avoid hybridization effects on the
 175 estimates. The admixture proportions were obtained from STRUCTURE.

176 The variance components across loci for hierarchical F-statistics for pure graylags and pure
 177 European domestics were estimated using locus-by-locus analysis of molecular variance (AMOVA)
 178 implemented in Arlequin 3.5.2.1 (Excoffier and Lischer 2010). The significance was tested with 16 000
 179 permutations.

180 e) Population structure analyses

181 Population clustering and structure was analyzed with STRUCTURE 2.3.4 (Pritchard *et al.* 2000) and
 182 Principal Component Analysis (PCA) (Patterson *et al.* 2006). For the whole dataset, STRUCTURE was
 183 run with 1000 burn-in steps followed by 10 000 iterations of MCMC for data collection for $K = 1-10$
 184 allowing admixture with five replicates of each run to reach convergence. For the STRUCTURE
 185 analyses done separately on graylags and European domestic geese, see Supplementary File 1. An
 186 admixture model with correlated allele frequencies among populations (Falush *et al.* 2003) was used
 187 in all STRUCTURE analyses and the iterations were automated with StrAuto 1.0 (Chhatre and Emerson
 188 2017). We applied both likelihood of K and Evanno's ΔK (Evanno *et al.* 2005) of successive K values to
 189 determine the optimal number of clusters, using STRUCTURE HARVESTER (Earl and VonHoldt 2012).
 190 CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) was used to align the assignments from different
 191 replicates of K and DISTRUCT 1.1 (Rosenberg 2003) for visualization. A PCA was performed with the
 192 prcomp function in R (R Core Team 2017) and the significance of eigenvalues determined based on
 193 the Tracy-Widom distribution (Patterson *et al.* 2006; van Heerwaarden *et al.* 2011).

194 A neighbor-joining tree was constructed for phylogenetic analysis, with pairwise distance
 195 between individuals obtained with the R package ape (Paradis *et al.* 2004) based on 40,191 loci. The
 196 *A. cygnoid* reference genome and the invariant sites that differed from it were included in the tree
 197 construction.

198 f) Tests for admixture and simulations of demographic history

199 The history of admixture was tested with a 3-Population test $f_3(C; A, B)$ implemented in AdmixTools
 200 4.1 (Patterson *et al.* 2012). This method offers a formal test to explain observed patterns of admixture
 201 in a target population without an outgroup. For identification of admixture between Chinese and
 202 European domestics, Grey and White Chinese were combined to represent the Chinese, and the
 203 Landes breed that had minimum indication of admixture in STRUCTURE was chosen to represent the
 204 European domestic source population. In addition, we tested several combinations of graylag geese,
 205 European domestic geese, and Chinese domestic geese as source populations to detect possible
 206 admixture in populations and breeds that implied admixture in STRUCTURE. See also Supplementary
 207 File 1 for further information.

208 Different models of demographic history were tested with fastsimcoal2 ver 2.6 (Excoffier *et al.*
 209 2013). Fastsimcoal2 uses coalescent simulations to estimate the likelihood of a demographic model
 210 and the probabilities obtained from simulations are then used to compute the composite likelihood of
 211 the model. The likelihood is maximized with a conditional maximization algorithm (ECM). We
 212 excluded all SNPs that had missing data within the whole data set and executed the analyses with a
 213 site frequency spectrum (SFS) based on 6,229 SNPs (Supplementary File 1, Figure S2). As there are no
 214 estimates of the genetic diversity per base pair for graylags, we estimated the proportions of variable
 215 and monomorphic sites in the data as we needed the information about the invariant sites for the
 216 fastsimcoal2 analysis. From the BAM file with `–depth` option in SAMtools 1.7 (Li *et al.* 2009), we
 217 estimated 9,801,382 bp covered with GBS tags. We then mimicked the filtering steps done for the
 218 biallelic SNPs to reduce the total number of sites in equivalent proportions. We removed the same
 219 number of sites that corresponded to the number of SNPs that were removed because they were
 220 indels, had more than 2 alleles or had heterozygosity over 0.75. Since some of the SNPs were

221 removed from this analysis due to missing data in some individuals, we removed an equal proportion
 222 of sites from the total number of sites as well. The final folded SFS had 1,681,316 sites of which
 223 1,675,087 were monomorphic and 6,229 polymorphic.

224 To infer the demographic history, we chose a subset of individuals from both wild-collected
 225 graylags and domestic geese to represent the genetic variation in both groups. Therefore, 11 graylags
 226 with > 90.8% of graylag ancestry and 15 domestic geese with > 91.4% of European domestic goose
 227 ancestry were selected for the analysis. The mutation rate for the simulations was $1.38 \cdot 10^{-7}$ per
 228 generation (Pujolar *et al.* 2018). The parameter estimation for each model tested involved 100,000
 229 simulations and 40 conditional maximization (ECM) cycles. The parameters for each model were
 230 estimated with 100 independent runs to obtain the global maximum. The models tested were i)
 231 simple divergence of two populations with no gene flow, ii) divergence of two populations with
 232 continuous gene flow and iii) divergence of two populations with changing gene flow patterns (Figure
 233 2, Figure S3-S4). The best model was selected based on Akaike's weight of evidence as in Excoffier *et*
 234 *al.* (2013). For parametric bootstrapping 100 SFS were simulated with the parameter estimates
 235 obtained from the real SFS, followed by maximum likelihood estimation with 50 independent runs for
 236 each bootstrap SFS. The 95% confidence intervals were obtained from the bootstrap data for each
 237 estimated parameter.

238 **Data availability**

239 The Supplementary File 1 that contains extended Materials and Methods, and Results including
 240 supplementary figures and tables, and Supplementary File 2 containing commands for the Tassel
 241 pipeline and vcftools are stored in figshare along with the VCF file containing the filtered genotypes.
 242 The raw sequence reads are available in NCBI's Sequence Read Archive (SRA) under BioProject
 243 PRJNA634849.

244 3. Results

245 a) Population structure

246 There was clear genetic differentiation between graylags and domestic geese according to
 247 STRUCTURE and PCA (Figure 3A-B). STRUCTURE aims to find the optimal number of ancestral
 248 populations (K) from the given data and the subdivision was clear in our data. At $K = 2$,
 249 populations/breeds are clustered based on their status (wild or domestic) and, at $K = 3$, domestic
 250 geese are further separated into European and Chinese. At $K = 4$, the fourth cluster is within graylag
 251 populations but none of the individuals are unanimously assigned to that cluster. The likelihood was
 252 highest for $K = 3$. These results were supported by PCA as the first two PCs out of 14 significant PCs (p
 253 < 0.05) were enough to separate the three groups (wild, European domestic, Chinese domestic) from
 254 each other (Figure 3A). Overall, the graylag populations showed 3.2% - 23.5% admixture proportions
 255 with European domestic geese when $K = 3$ (Table S1). In contrast, not all European domestic geese
 256 showed admixture with graylags and the admixture percentages ranged from 0.0 to 8.4%. At $K = 3$
 257 many European domestic goose breeds showed mixed ancestry with Chinese domestic geese (0.0 -
 258 27.1%).

259 The neighbor-joining tree repeated the major patterns observed with STRUCTURE and PCA,
 260 revealing a star shaped phylogeny and confirming that the domestic and graylag geese largely form
 261 different clades (Figure S5). Surprisingly, the Chinese domestic geese were closer to European
 262 domestic geese and graylags, than to the swan goose reference genome. In addition, one graylag
 263 from Turkey was more closely related to the Chinese domestic geese than other graylags, also
 264 indicated by admixture proportions from STRUCTURE. Further, two Crested Faroese individuals and
 265 four domestics from the USA (2 unknown and 2 Toulouse crosses) were closer to Chinese than

European domestic geese. These six individuals also showed high proportions of admixture with Chinese domestics in the STRUCTURE analysis.

Unequal sample sizes did not have a large effect on the results (Supplementary File 1, Figure S6-S11). Some further population structure was observed within both graylags and domestic geese, when analyzed separately with STRUCTURE and PCA. Geographically, graylags differentiated by subspecies (Supplementary File 1, Figure S12-S13). STRUCTURE indicated little differentiation among European domestic geese, but the PCA revealed separation between the European breeds and the Turkish domestic geese (Supplementary File 1, Figure S14-S15).

b) Genetic diversity

An AMOVA was used to partition genetic diversity among graylag vs. domestic (group level), and among populations (graylag) and among breeds (domestic), and within population levels (Table 1). The fixation index between graylag and domestic geese was 0.158 and there was also significant differentiation among graylag populations/domestic breeds (Table 1). The average pairwise F_{ST} between graylag populations and domestic breeds was 0.197, among graylag populations 0.088 and among domestic breeds 0.174 (Supplementary File 1, Table S2).

The genetic diversity measured as average H_E was higher in pure graylags (0.146) than in pure European domestic geese (0.096) (Welch's t-test, degrees of freedom (df) = 10.594, $p = 3.91 \times 10^{-5}$, see also Supplementary File 1, Figure S16). The average H_E ranged from 0.140 (Denmark) to 0.150 (Kazakhstan) in pure graylags and from 0.047 (Landes) to 0.123 (Domestic N-Turkey) in pure European domestics. The difference in average H_E remained when non-pure graylag and non-pure European domestics were included in the comparison (0.156 vs. 0.107; Welch's t-test, df = 19.28, $p = 0.000418$). The average H_E was higher in admixed populations compared to non-admixed populations in both graylag and domestic populations (Supplementary File 1; Table S1, Figure S16).

289 c) Admixture and the time of domestication

290 STRUCTURE implied considerable mixed ancestry from multiple genetic clusters for Dutch and Turkish
 291 graylags, but the f_3 analysis did not confirm admixture for the Dutch population even though multiple
 292 source populations of graylag and domestic goose were tested (Table S3). However, the Turkish
 293 population is more complicated as they obtained negative f_3 when analyzed together with multiple
 294 combinations of source populations indicating admixture with Chinese domestic goose but not with
 295 European domestic goose. This signal appeared consistently when several graylag and European
 296 domestic goose populations were used as source populations with Chinese domestic geese. However,
 297 as the Turkish graylags appeared genetically very dissimilar, we analyzed them separately which
 298 resulted in neither of them obtaining negative f_3 (Table S3). The two Turkish graylag samples came
 299 from the same area as our NW-Turkish domestic population, which among Turkish domestic geese
 300 showed highest admixture with graylags (2.2%), but admixture was not confirmed with the f_3 test
 301 (Table S4). We did not obtain negative Z-scores to any of the other graylag populations either (Table
 302 S5-S6).

303 The f_3 analysis confirmed admixture of domestic geese in line with the STRUCTURE results. Most
 304 notably, the African breed is a hybrid between European and Chinese domestic geese (Z-score -
 305 6.399), unexpected as this breed has been assumed to have originated solely from swan goose. The
 306 European-Chinese hybrid status of the Kholmogory and Steinbacher breeds was also confirmed (Z-
 307 scores of -8.933 and -5.349, respectively). The Kholmogory breed also fell halfway between European
 308 and domestic geese both in STRUCTURE and PCA, whereas the Steinbacher was genetically closer to
 309 European domestic geese in the PCA. However, the Diepholzer breed, which reportedly is also a
 310 hybrid, was not confirmed as such in our analysis. Other domestic breeds/groups with admixture
 311 status in STRUCTURE were also confirmed to have a European-Chinese admixture when a Z-score

threshold of -3 (roughly corresponding to $p < 0.01$) was used: Sebastopol, Toulouse cross, Domestic NY, Embden, Tufted Roman (Figure 3C, Supplementary File 1, Table S5). These breeds also gave a similar signal when other combinations of European domestic goose breeds and Chinese domestic geese were used as source populations (Table S7). The Crested Faroese breed gave indication of admixture based on STRUCTURE analysis and the f_3 test supported this (Z-score of -2.228, $p < 0.05$). Surprisingly, the Northern Turkish domestic population was not admixed with Chinese domestic geese in STRUCTURE but f_3 analysis gave a contrasting signal (Z-score -2.459, $p < 0.05$).

The demographic model that best fit our data suggested divergence of graylag and domestic geese with a recent migration rate change (Table 2, Supplementary File 1, Table S8). The model suggested divergence around 5319 generations ago (95% confidence intervals (CI): 2014-6503) with asymmetric but close to equal migration rates from graylags to domestic geese following divergence. About 159 (88-476) generations ago, there was a change in the gene flow patterns, suggesting higher gene flow (m) from graylag geese to domestic geese towards modern times. However, translated to actual number of migrants ($N_e m$), the numbers suggest that the gene flow has been higher from domestic geese to graylag geese across domestication history, (0.41 graylag geese vs. 1.34 domestic geese migrating per generation following the domestication event, and 1.65 graylag geese vs. 1.67 domestic geese per generation migrating after the gene flow pattern changed). Given an estimated generation time for these geese of about 3 years, the numbers suggest divergence about 14 000 BCE and gene flow shift about 480 years ago.

4. Discussion

We studied the dynamics of domestication and hybridization in grey (*Anser*) geese using genome-wide SNP data. The results demonstrated genetic divergence between Eurasian wild graylag and European domestic geese with long-term genetic exchange between them. We also inferred temporal

335 changes in the direction of gene flow. The degree of hybridization between graylag and domestic
 336 geese also varied geographically. Surprisingly, several domestic goose breeds also showed a
 337 substantial genetic contribution of Chinese domestic geese. We also provide insights about the origin
 338 and the timing of goose domestication.

339 a) Genetic diversity and differentiation of graylag and European domestic geese

340 Domestic species often show reduced genetic diversity compared to their wild ancestor, attributable
 341 to genetic drift during population bottlenecks of initial domestication, combined with subsequent
 342 artificial selection associated with breed formation (Moyers *et al.* 2018). Domestic geese appear to
 343 follow the same trend. We found European domestic geese to have lower H_E than wild graylags. In
 344 general, graylag populations were much more uniform in their level of genetic diversity whereas
 345 domestic populations showed more variance, which is likely to reflect the human influence on breed
 346 formation.

347 European domestic geese are genetically distinct from their wild progenitor but no more so
 348 than for other domestic birds. The average pairwise F_{ST} values between graylag populations and
 349 domestic goose breeds were lower than between red junglefowl and domestic chicken populations
 350 (Kanginakudru *et al.* 2008), and domestic geese are less distinctive than domestic pigeons (Stringham
 351 *et al.* 2012). Among domestic geese, the Turkish are particularly interesting. From mtDNA, the Turkish
 352 domestic geese stand out as the most genetically variable group (Heikkinen *et al.* 2015), and although
 353 this is less evident from GBS, among the pure European domestic geese the Northern Turkish showed
 354 the highest average H_E . The f_3 analysis indicates a history of admixture with Chinese domestics for this
 355 population, which may explain its high genetic diversity.

356 We found a genetic separation between European and Near Eastern populations of graylags
 357 that aligned with the western and eastern subspecies (*A. a. anser* and *A. a. rubrirostris*) (Scott and

358 Rose 1996), a distinction which could not be made based on mtDNA (Heikkinen *et al.* 2015).
 359 Hybridization between the western and eastern subspecies is suggested from admixture in Dutch and
 360 Danish graylags in STRUCTURE as there is a genetic component that is more prevalent in the eastern
 361 populations. There is historical evidence for the introduction of *rubrirostris* to Belgium in 1954 and to
 362 Netherlands in 1960s (Rooth 1971; Kuijken and Devos 1996); thus, *rubrirostris* genes may have
 363 originated from the recently introduced gene pool spreading to Denmark.

364 b) When and where were geese domesticated?

365 Traditional views on goose domestication claim it first occurred in the eastern Mediterranean
 366 (possibly Egypt) around the 3rd Millennium BCE (Zeuner 1963; Albarella 2005). Domestication of
 367 chicken and perhaps pigeon took place earlier, but domestication of duck later, at least in Europe
 368 (Larson and Fuller 2014). Demographic modelling suggests that the wild graylag and related domestic
 369 lineages split approximately 5,300 generations ago placing domestication origins at 14 000 BCE
 370 assuming a 3-year generation time (Cramp and Simmons 1977). This estimated genetic divergence
 371 time is, admittedly, considerably earlier than any evidence for animal domestication except dog. It is
 372 important to note that the estimated divergence times have large confidence intervals and merely
 373 indicates the split between the ancestors of contemporary wild and domestic lineages. It is most likely
 374 that our demographic modelling reflects the early divergence of different lineages of graylags, only
 375 one of which contributed to later domestication. The subsequent reduction or even disappearance of
 376 that wild lineage means that, despite wide geographical sampling, the possible modern wild
 377 population(s) of the graylag progenitor to domestic geese was not sampled in this study. It is also
 378 worth remembering that using *A. cygnoid* reference genome may have caused a mapping bias of *A.*
 379 *anser* alleles failing to map on the reference genome due to sequence divergence. This would have
 380 affected the subsequent SNP calling by reducing the number of rare, derived *A. anser* alleles, which in

381 turn could cause our divergence time estimate to be an underestimate. Another thing to bear in mind
 382 is the uncertainty about the mutation rate. The estimate we used by Pujolar *et al.* (2018) was
 383 estimated for pink-footed goose which is a closely related to graylag goose and was supported by
 384 Ottenburghs *et al.* (2016b) who obtained a similar substitution rate for geese. However, both
 385 estimates are about two orders of magnitude higher than that estimated for collared flycatcher using
 386 pedigree data (Smeds *et al.* 2016). It is possible that this is a taxon-related difference but in case the
 387 substitution rate for graylag goose is actually closer to that of collared flycatcher, the mutation rate
 388 we used here would be too high and our estimate of the domestication time would have to be pushed
 389 even further back. Therefore, the estimated divergence time should be considered as a guideline for
 390 future studies and not as an absolute truth. Future studies would benefit from whole genome
 391 sequencing of graylag goose in resolving the questions about both mapping bias and the substitution
 392 rate.

393 Given that genetic diversity would be expected to be highest in the ‘domestication center’ and
 394 reduce with increasing distance from there, the high mtDNA diversity of Turkish domestic geese
 395 means the eastern Mediterranean cannot be ruled out as a candidate for the origin of goose
 396 domestication. However, as we have shown, hybridization between wild and domestic geese can also
 397 generate high genetic diversity both within and outside the original domestication location. More
 398 thorough sampling of the graylag population around the Black Sea would be beneficial in resolving the
 399 role of eastern Mediterranean region in the domestication history of goose as this population was not
 400 well represented in our study. Additionally, the progenitor of domestic geese could be sought by
 401 ancient DNA approaches.

402 c) The role of intra- and interspecific hybridization in goose domestication history

403 i. Evidence of current hybridization

Domestic animals and their wild relatives are often observed to interbreed, and this is also true for geese. Both field observations and mtDNA results (Kuijken and Devos 1996; Heikkinen *et al.* 2015) suggested some current hybridization between domestic and graylag geese. Genome-wide analysis covering multiple graylag populations and domestic breeds revealed a considerable impact of hybridization on genetic diversity of both wild and domestic geese.

Hybridization is particularly prevalent in certain geographical regions. Dutch and especially Turkish wild graylag samples had more shared genetic affiliation with domestics than Scandinavian and Finnish graylag populations (Figure 3B). Some regions may offer more hybridization opportunities, e.g. climate may allow graylags to be sedentary year-round and be favorable for keeping domestic geese. The Netherlands, for instance, lies on the Atlantic flyway offering breeding, staging, and wintering areas for graylags (Madsen *et al.* 1999; Andersson *et al.* 2001). Since pair-bonding of geese generally occurs on wintering grounds (Rohwer and Anderson 1988), hubs for migrating geese such as the Netherlands may permit population mingling. Nevertheless, the f_3 test did not support a simple history of admixture for the Netherlands. Patterson *et al.* (2012) have stated that population-specific drift may mask the signal of admixture in such analyses, leading to a non-negative f_3 . The f_3 model is relatively simple, with only two sources, and may not catch the signal of admixture in the Dutch graylag population because of the previous contribution of *rubrirostris*, which was not included in the model.

Based on ringing data most graylag populations in Scandinavia follow the Atlantic flyway - some of the geese wintering in the Netherlands and others in southwest Spain. However, Finnish graylags favor the Central European flyway and winter in North Africa, with a minority of Finnish graylags using the Atlantic Flyway (Madsen *et al.* 1999; Andersson *et al.* 2001). The Finnish populations of graylag showed the lowest admixture proportions with domestic geese (S-Finland 3.2% domestic goose, N-

427 Finland 3.3% domestic geese) among graylag populations. Rearing geese is not a popular practice in
 428 Finland, and they constituted less than 5% of poultry kept in Finland in 2014 (“Official Statistics of
 429 Finland (OSF): Number of livestock [e-publication]. Helsinki: Natural Resources Institute Finland
 430 [referred: 17.12.2016]. Access method: http://www.stat.fi/til/klm/index_en.html” 2016). The
 431 Norwegian populations showed only slightly higher admixture proportions with domestic geese,
 432 although the domestic mtDNA haplotype ANS19 was detected from a wild graylag collected in
 433 Finnmark, Norway (Pellegrino *et al.* 2015). This haplotype is a partial sequence of the D5 haplotype
 434 identified by Heikkinen *et al.* (2015), and identical to that found in White Roman domestic geese
 435 (Wang *et al.* 2010).

436 Inferring the hybridization patterns in the Turkish graylags is more complicated, as Turkish
 437 graylags indicate hybridization with both Chinese and European domestics. Both graylags sampled in
 438 Turkey showed considerable admixture with domestic geese. One of them appeared genetically as a
 439 hybrid of European and Chinese domestic goose with only a small proportion of graylag ancestry,
 440 whereas the other one was a more equal mix of European domestic goose and graylag supplemented
 441 by a considerable Chinese domestic goose ancestry. However, what appears as a hybridization
 442 between European and Chinese domestic geese may also be related to ancestral variation, and result
 443 from close relatedness of the Turkish graylags to the graylag population that was domesticated,
 444 reinforced by a gene flow from the Chinese domestic goose. There is some indication of hybridization
 445 between graylags and domestic geese within that area as the domestic geese sampled from the same
 446 area showed some admixture with graylags, but this was not confirmed with f_3 analysis. These results
 447 may reflect a local practice of keeping captive graylags within a flock of domestic geese as several
 448 sources state that it has been a common practice to collect wild eggs and goslings in many places
 449 across Eurasia (Gray 1871; Honka *et al.* 2018). Another possibility is that the Turkish graylags have

450 hybridized with some unsampled distinct graylag population and simply appear genetically like
451 domestic geese due to lack of representation of the unsampled wild population. The graylag
452 population breeding and wintering in the Black Sea region is not well monitored (Fox *et al.* 2010).

453 ii. Long-term hybridization

454 Domestication can be seen as an analogy of speciation where an animal population transforms to an
455 ecotype that is adapted to the human niche (Larson and Fuller 2014) and at later stages of
456 domestication is perpetuated with reproductive isolation in the form of selection managed by
457 humans (Zeder 2012). However, this reproductive isolation may not be complete (Frantz *et al.* 2015).
458 While the genetic divergence of the graylag and its domestic descendant is evident, our results
459 suggest extensive long-term genetic exchange between them. In addition, the demographic modelling
460 suggests that the gene flow patterns have changed over time.

461 Initially, gene flow was greater from domestic geese to graylag geese. It is unlikely that the early
462 stages of goose domestication were rigorously managed, allowing matings outside the domestic gene
463 pool. It is in the farmers' interest to keep the domestic geese and wild geese reproductively isolated
464 to keep control over the traits that are being selected, but artificial selection of traits would have
465 become possible only after the domestic gene pool had been established. After that, it may
466 occasionally be beneficial to restock the flock to maintain enough genetic diversity. Several sources
467 have suggested that it has been a common practice to collect goose eggs from the wild and raise
468 them in captivity. The natural tendency for imprinting in geese facilitates this practice. Goose-keeping
469 became well-established in the Medieval period (Albarella 2005) and the rise in number of domestic
470 geese may have allowed an increase in domestic goose escapees resulting in increased gene flow
471 (N_{em}) from domestic geese back to graylags towards modern times.

Furthermore, not only have domestic geese admixed with wild graylags but also European and Chinese domestic geese have hybridized. Hybridization with ancestral species or closely related species is frequent in domestic species, e.g., the genetic composition of chicken derives from multiple different species of *Gallus* (Eriksson *et al.* 2008). Similarly, the genetic composition of domestic geese seems to derive from two closely related species. This hybridization with Chinese domestic geese may have introduced some traits not present in graylags to European domestic geese and vice versa.

5. Conclusion

This study is the first attempt to answer questions related to goose domestication history using population genetic approach with genome-wide data. We have shown that hybridization has played and continues to play a significant role in shaping the wild and domestic graylag populations. Admittedly, the demographic models we used here were quite simple and they are unlikely to capture every nuance of the population history, but they offer a starting point for future studies which may include more elaborate analyses of demographic history, for example changes in effective population size associated with population bottlenecks during domestication. Selection scans could be used to identify introgressed alleles that have been under selection during domestication. The use of whole genome sequencing would be advantageous in aforementioned analyses and would also enable assessment of runs of homozygosity (ROH) in goose genome.

Authors' contributions.

MEH conceived the study, contributed to data collection, analyzed the data, and drafted the manuscript. MR acquired the funding, conceived the study, and contributed to data collection. TAW contributed to data collection and participated to data analysis. MMA and IG contributed to data collection. KMD, JA, JBS and TP conceived the study and contributed to writing and interpretation of

494 data, with TP also participating in data analysis. All authors, excluding MR, reviewed, improved, and
495 approved the manuscript.

496 **Competing interests.**

497 The authors declare no conflict of interest.

498 **Funding.**

499 This work was supported by the Academy of Finland (grant no. 131673 to MR, grant no 283609 to JA
500 and grant no. 287431 to TP); and the Cornell Center for Comparative and Population Genomics (3CPG)
501 to JBS. The Emil Aaltonen Foundation and Oulun läänin talousseuran maataloussäätiö are
502 acknowledged for the personal grants awarded to MEH.

503 **Acknowledgements.**

504 We thank Jon Fjeldså, Abolghasem Khaleghizadeh, Ellen Vuosalo-Tavakoli, Sofia Rozenfield, Maarten
505 Loonen, Arne Follestad, Annita Logotheti and Natural Resources Institute Finland for their help in
506 obtaining the graylag goose samples. We also thank Svetlana Pavlova, Helle Palmø from Ministeriet
507 for Fødevarer, Landbrug og Fiskeri in Denmark, Mikael Olsson from Svenska Lanthönsklubben in
508 Sweden, and The Goose Club in the UK for their help in obtaining the domestic goose samples. Many
509 thanks to the goose breeders in Denmark, Sweden, Turkey, UK and USA who provided the domestic
510 goose samples. Silvia Markova is acknowledged for her help with the laboratory work. We are grateful
511 to Tiina Mattila for their help with the fastsimcoal2 analysis. We thank the two anonymous reviewers
512 for their comments to improve the manuscript. Cornell Biotechnology Resource Center is
513 acknowledged for the GBS library construction and sequencing. Lastly, we thank the CSC – IT Center
514 for Science in Finland for providing computing resources.

515 **References**

- 516 Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird *et al.*, 2013 Hybridization and speciation. J.
517 Evol. Biol. 26: 229–246.
- 518 Albarella, U., 2005 Alternate fortunes? The role of domestic ducks and geese from Roman to Medieval
519 times in Britain, pp. 249–258 in *Documenta Archaeobiologiae 3. Feathers, grit and symbolism;*
520 *Birds and humans in the ancient Old and New Worlds.*, edited by G. Grupe and J. Peters. Rahden:
521 Verlag Marie Leidorf.
- 522 Andersson, Å., A. Follestad, L. Nilsson, and H. Persson, 2001 Migration patterns of Nordic greylag
523 geese *Anser anser*. Ornis Svecica 11: 19–58.
- 524 Arnold, M. L., 2004 Natural hybridization and the evolution of domesticated, pest and disease
525 organisms. Mol. Ecol. 13: 997–1007.
- 526 BirdLife International and Handbook of the Birds of the World (2016) 2016. *Anser anser*. The IUCN Red
527 List of Threatened Species. Version 2018-1.
- 528 Buckland, R., and G. Guy (Eds.), 2002 Goose production, in *FAO Animal Production and Health Paper -*
529 *154*, Food and Agriculture Organization of the United Nations, Rome.
- 530 Chhatre, V. E., and K. J. Emerson, 2017 StrAuto: automation and parallelization of STRUCTURE
531 analysis. BMC Bioinformatics 18: 192.
- 532 Cramp, S., and K. E. L. Simmons (Eds.), 1977 *Handbook of the birds of Europe, the Middle East, and*
533 *North Africa: The birds of the Western Palearctic. Vol. 1: Ostrich-Ducks*. Oxford University Press,
534 New York.
- 535 Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and

- 536 VCFtools. *Bioinformatics* 27: 2156–2158.
- 537 Delacour, J., 1954 *The waterfowl of the world : Vol. 1, The magpie goose, whistling ducks, swans and*
 538 *geese, sheldgeese and shelducks*. Country Life, London.
- 539 Earl, D. A., and B. M. VonHoldt, 2012 STRUCTURE HARVESTER: a website and program for visualizing
 540 STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4: 359–361.
- 541 Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-
 542 by-sequencing (GBS) approach for high diversity species. *PLoS One* 6: e19379.
- 543 Eriksson, J., G. Larson, U. Gunnarsson, B. Bed’hom, M. Tixier-Boichard *et al.*, 2008 Identification of the
 544 yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet.* 4: e1000010.
- 545 Evanno, G., S. Regnaut, and J. Goudet, 2005 Detecting the number of clusters of individuals using the
 546 software structure: a simulation study. *Mol. Ecol.* 14: 2611–2620.
- 547 Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll, 2013 Robust demographic
 548 inference from genomic and SNP data. *PLoS Genet.* 9: e1003905.
- 549 Excoffier, L., and H. E. L. Lischer, 2010 Arlequin suite ver 3.5: a new series of programs to perform
 550 population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10: 564–567.
- 551 Falush, D., M. Stephens, and J. K. Pritchard, 2003 Inference of population structure using multilocus
 552 genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- 553 Fox, A. D., B. S. Ebinger, C. Mitchell, T. Heinicke, T. Aarvak *et al.*, 2010 Current estimates of goose
 554 population sizes in western Europe, a gap analysis and an assessment of trends. *Ornis Svecica* 20:
 555 115–127.
- 556 Frantz, L. A. F., J. G. Schraiber, O. Madsen, H.-J. Megens, A. Cagan *et al.*, 2015 Evidence of long-term

- 557 gene flow and selection during domestication from analyses of Eurasian wild and domestic pig
558 genomes. *Nat. Genet.* 47: 1141–1148.
- 559 Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: a high capacity
560 genotyping by sequencing analysis pipeline. *PLoS One* 9: e90346.
- 561 Godinho, R., L. Llana, J. C. Blanco, S. Lopes, F. Álvares *et al.*, 2011 Genetic evidence for multiple
562 events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Mol. Ecol.*
563 20: 5154–5166.
- 564 Goudet, J., 2005 hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol.*
565 Notes 5: 184–186.
- 566 Grant, P. R., and B. R. Grant, 1992 Hybridization of bird species. *Science* 256: 193–197.
- 567 Gray, R., 1871 *The Birds of the West of Scotland: Including the Outer Hebrides, with Occasional*
568 *Records of the Occurrence of the Rarer Species Throughout Scotland Generally*. T. Murray & Son,
569 Glasgow.
- 570 van Heerwaarden, J., J. Doebley, W. H. Briggs, J. C. Glaubitz, M. M. Goodman *et al.*, 2011 Genetic
571 signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad.*
572 *Sci. USA* 108: 1088–1092.
- 573 Heikkinen, M. E., M. Ruokonen, M. Alexander, J. Aspi, T. Pyhäjärvi *et al.*, 2015 Relationship between
574 wild greylag and European domestic geese based on mitochondrial DNA. *Anim. Genet.* 46: 485–
575 497.
- 576 Honka, J., M. Heino, L. Kvist, I. Askeyev, D. Shaymuratova *et al.*, 2018 Over a thousand years of
577 evolutionary history of domestic geese from Russian archaeological sites, analysed using ancient
578 DNA. *Genes* 9: 367.

- 579 Hufford, M. B., P. Lubinsky, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand *et al.*, 2013 The genomic
580 signature of crop-wild introgression in maize. *PLoS Genet.* 9: e1003477.
- 581 Jakobsson, M., and N. A. Rosenberg, 2007 CLUMPP: a cluster matching and permutation program for
582 dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*
583 23: 1801–1806.
- 584 Janzen, G. M., L. Wang, and M. B. Hufford, 2019 The extent of adaptive wild introgression in crops.
585 *New Phytol.* 221: 1279–1288.
- 586 Kanginakudru, S., M. Metta, R. Jakati, and J. Nagaraju, 2008 Genetic evidence from Indian red jungle
587 fowl corroborates multiple domestication of modern day chicken. *BMC Evol. Biol.* 8: 174.
- 588 Kear, J., 1990 *Man and wildfowl*. T & AD Poyser Ltd, London.
- 589 Kuijken, E., and K. Devos, 1996 The status of the greylag goose *Anser anser* in Flanders, Belgium. *Wetl.*
590 *Int. Goose Spec. Gr. Bull.* 8: 26–28.
- 591 Larson, G., and D. Q. Fuller, 2014 The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.*
592 45: 115–136.
- 593 Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform.
594 *Bioinformatics* 25: 1754–1760.
- 595 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format
596 and SAMtools. *Bioinformatics* 25: 2078–2079.
- 597 Lu, L., Y. Chen, Z. Wang, X. Li, W. Chen *et al.*, 2015 The goose genome sequence leads to insights into
598 the evolution of waterfowl and susceptibility to fatty liver. *Genome Biol.* 16: 89.
- 599 Madsen, J., G. Cracknell, and T. Fox (Eds.), 1999 *Goose populations of the Western Palearctic. A review*

- 600 *of status and distribution*. Wetlands International Publication No. 48, Wageningen, Ronde.
- 601 Marshall, F. B., K. Dobney, T. Denham, and J. M. Capriles, 2014 Evaluating the roles of directed
602 breeding and gene flow in animal domestication. *Proc. Natl. Acad. Sci. USA* 111: 6153–6158.
- 603 McTavish, E. J., J. E. Decker, R. D. Schnabel, J. F. Taylor, and D. M. Hillis, 2013 New World cattle show
604 ancestry from multiple independent domestication events. *Proc. Natl. Acad. Sci. USA* 110:
605 E1398–E1406.
- 606 Moyers, B. T., P. L. Morrell, and J. K. McKay, 2018 Genetic costs of domestication and improvement. *J.*
607 *Hered.* 109: 103–116.
- 608 Official Statistics of Finland (OSF): Number of livestock [e-publication]. Helsinki: Natural Resources
609 Institute Finland [referred: 17.12.2016]. Access method:
610 http://www.stat.fi/til/klm/index_en.html, 2016.
- 611 Ottenburghs, J., P. van Hooft, S. E. van Wieren, R. C. Ydenberg, and H. H. T. Prins, 2016a Hybridization
612 in geese: a review. *Front. Zool.* 13: 20.
- 613 Ottenburghs, J., H.-J. Megens, R. H. S. Kraus, O. Madsen, P. van Hooft *et al.*, 2016b A tree of geese: a
614 phylogenomic perspective on the evolutionary history of True Geese. *Mol. Phylogenet. Evol.* 101:
615 303–313.
- 616 Paradis, E., J. Claude, and K. Strimmer, 2004 APE: analyses of phylogenetics and evolution in R
617 language. *Bioinformatics* 20: 289–290.
- 618 Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland *et al.*, 2012 Ancient admixture in human
619 history. *Genetics* 192: 1065–1093.
- 620 Patterson, N., A. L. Price, and D. Reich, 2006 Population structure and eigenanalysis. *PLoS Genet.* 2:
621 e190.

- 622 Pellegrino, I., M. Cucco, A. Follestad, and M. Boos, 2015 Lack of genetic structure in greylag goose
 623 (*Anser anser*) populations along the European Atlantic flyway. *PeerJ* 3: e1161.
- 624 Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus
 625 genotype data. *Genetics* 155: 945–959.
- 626 Pujolar, J. M., L. Dalén, R. A. Olsen, M. M. Hansen, and J. Madsen, 2018 First de novo whole genome
 627 sequencing and assembly of the pink-footed goose. *Genomics* 110: 75–79.
- 628 R Core Team, 2017 R: A Language and Environment for Statistical Computing. [https://www.R-](https://www.R-project.org/)
 629 [project.org/](https://www.R-project.org/)
- 630 Ren, T., S. Liang, A. Zhao, and K. He, 2016 Analysis of the complete mitochondrial genome of the
 631 Zhedong White goose and characterization of NUMTs: reveal domestication history of goose in
 632 China and Euro. *Gene* 577: 75–81.
- 633 Rohwer, F. C., and M. G. Anderson, 1988 Female-biased philopatry, monogamy, and the timing of pair
 634 formation in migratory waterfowl, pp. 187–221 in *Current Ornithology*, edited by R. F. Johnston.
 635 Springer US, Boston, MA.
- 636 Rooth, J., 1971 The occurrence of the greylag goose *Anser anser* in the western part of its distribution
 637 area. *Ardea* 59: 17–27.
- 638 Rosenberg, N. A., 2003 distruct: a program for the graphical display of population structure. *Mol. Ecol.*
 639 *Notes* 4: 137–138.
- 640 Scott, D. A., and P. M. Rose, 1996 *Atlas of Anatidae populations in Africa and Western Eurasia*.
 641 Wetlands International Publication No 41, Wageningen.
- 642 Shi, X.-W., J.-W. Wang, F.-T. Zeng, and X.-P. Qiu, 2006 Mitochondrial DNA cleavage patterns
 643 distinguish independent origin of Chinese domestic geese and Western domestic geese.

- 644 Biochem. Genet. 44: 237–245.
- 645 Skoglund, P., E. Ersmark, E. Palkopoulou, and L. Dalén, 2015 Ancient wolf genome reveals an early
646 divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr. Biol. 25:
647 1515–1519.
- 648 Smeds, L., A. Qvarnström, and H. Ellegren, 2016 Direct estimate of the rate of germline mutation in a
649 bird. Genome Res. 26: 1211–1218.
- 650 Stringham, S. A., E. E. Mulroy, J. Xing, D. Record, M. W. Guernsey *et al.*, 2012 Divergence,
651 convergence, and the ancestry of feral populations in the domestic rock pigeon. Curr. Biol. 22:
652 302–308.
- 653 Sun, J., S. Zhang, D.-Q. He, S.-Y. Chen, Z.-Y. Duan *et al.*, 2013 Matrilineal genetic structure of domestic
654 geese. J. Poult. Sci. 51: 130–137.
- 655 Wang, C. M., T. D. Way, Y. C. Chang, N. T. Yen, C. L. Hu *et al.*, 2010 The origin of the White Roman
656 goose. Biochem. Genet. 48: 938–943.
- 657 Welch, B. L., 1938 The significance of the difference between two means when the population
658 variances are unequal. Biometrika 29: 350.
- 659 Zeder, M. A., 2012 The domestication of animals. J. Anthropol. Res. 68: 161–190.
- 660 Zeuner, F. E., 1963 *A History of Domesticated Animals*. Harper & Row, Publishers, New York and
661 Evanston.

662

663

664

665 **Table 1.** Hierarchical analysis of molecular variance (AMOVA) of graylags and their domestic descendants,
 666 considering pure populations of graylags (first group) and pure breeds of European domestic geese (second
 667 group).

SOURCE OF VARIATION	SUM OF	VARIANCE	PERCENTAGE	FIXATION
	SQUARES	COMPONENTS	VARIATION	INDICES
Among groups	47565.119	431.4291	15.8	$F_{CT} = 0.158^{**}$
Among populations and breeds	82960.489	302.51404	11.1	$F_{SC} = 0.131^{**}$
within groups				
Within populations and breeds	345889.821	2003.45893	73.2	$F_{ST} = 0.268^{**}$
Total	476415.429	2737.40207		

$^{**} p < 0.001$

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686 **Table 2.** Maximum likelihood estimates (MLE) for the parameters of the preferred demographic model for goose
 687 domestication history (see text) with their 95% confidence intervals (CI).

MODEL	PARAMETER	MLE	95% CI
Divergence with changing gene flow patterns	ANCSIZE	1112	378.95 - 7990.65
	T ₁	5319	2014.45 - 6503.75
	M1 _{WD}	4.25x10 ⁻⁴	1.21x10 ⁻⁷ - 6.28x10 ⁻⁴
	M1 _{DW}	5.35x10 ⁻⁴	2.88x10 ⁻⁴ - 6.45x10 ⁻⁴
	T ₂	159	88.9 - 476.25
	M2 _{WD}	1.72x10 ⁻³	1.30x10 ⁻³ - 2.23x10 ⁻³
	M2 _{DW}	6.69x10 ⁻⁴	4.17x10 ⁻⁴ - 8.00x10 ⁻⁴
	N _{WILD}	2504	2352.4 - 2680.25
	N _{DOM}	959	833.95 - 1040.55

ANCSIZE, effective population size of ancestral population; T₁, time of divergence in generations;
 N_{DOM}, effective population size for domestic geese; N_{WILD}, effective population size for graylags; T₂,
 estimate of time in generations when the migration matrix switched; M1_{WD} migration rate from wild to
 domestic following T₁; M1_{DW} migration rate from domestic to wild following T₁; M2_{WD} migration rate
 from wild to domestic following T₂; M2_{DW} migration rate from domestic to wild following T₂.

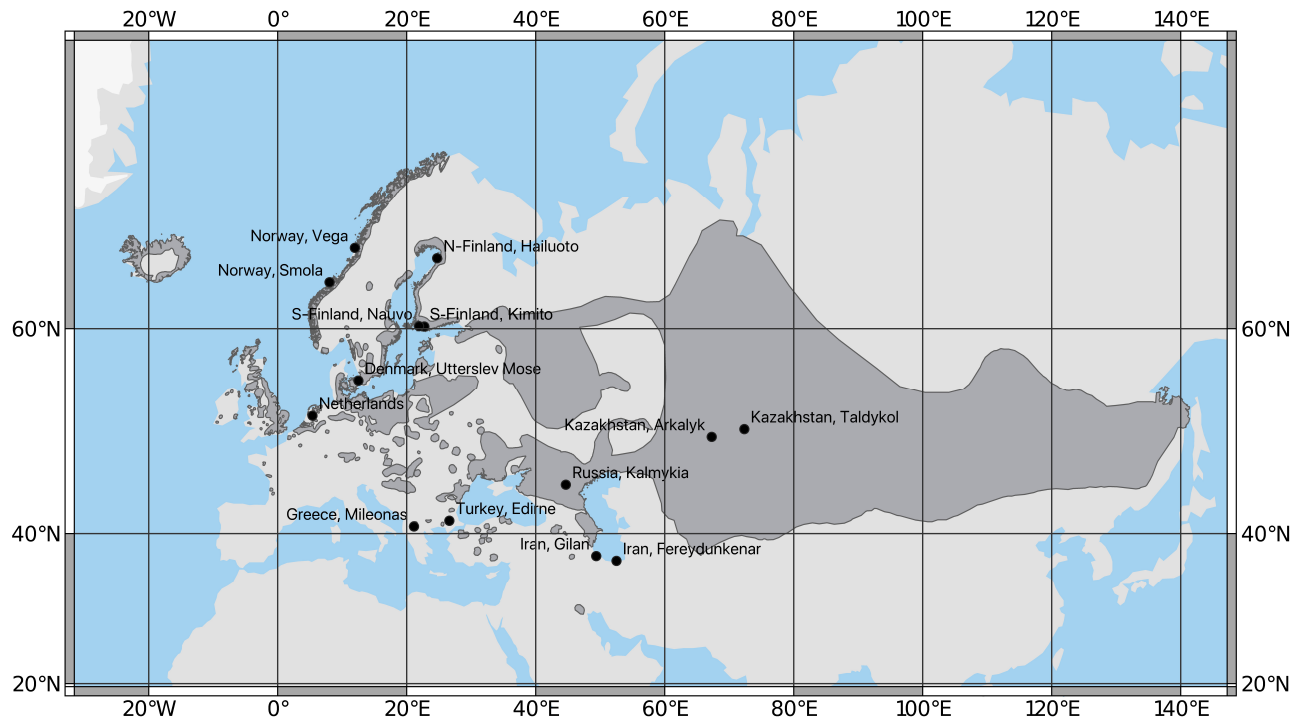


Figure 1. Map showing the sampling sites for wild graylags used in this study. The breeding area of the species is shown on darker grey. The sampling sites in Kazakhstan were combined for analyses (one sample per location) and the sampling sites in Southern Finland included combined samples from the geographically close sites of Västana fjärd, Nauvo (shown) and Kimito (shown). The Iranian samples were collected during the wintering season. Map modified from IUCN ("BirdLife International and Handbook of the Birds of the World (2016) 2016. *Anser anser*. The IUCN Red List of Threatened Species. Version 2018-1").

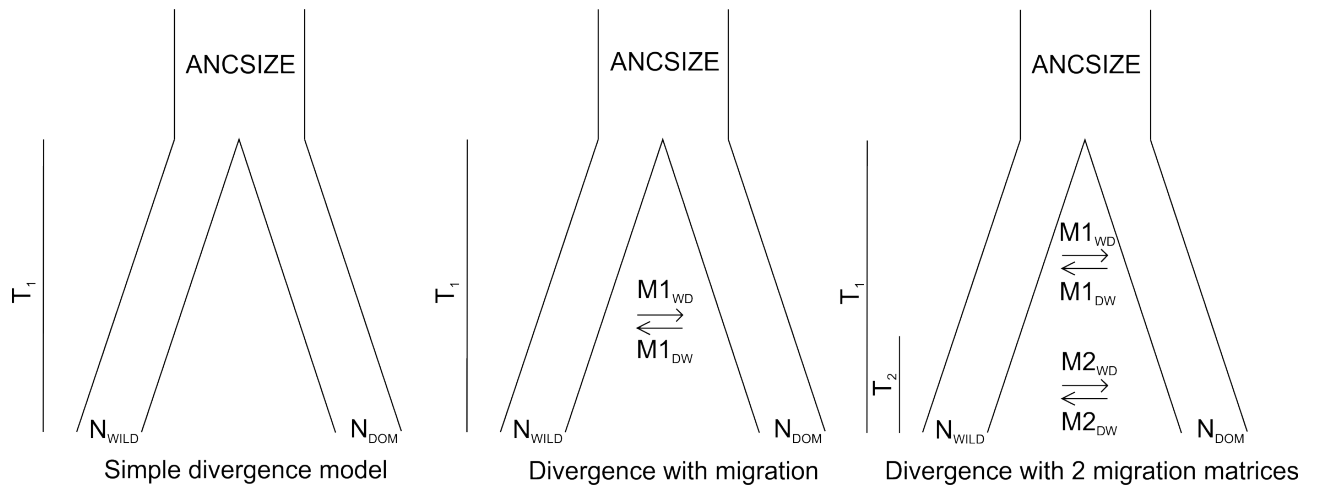


Figure 2. Demographic histories of goose domestication as tested with fastsimcoal2.

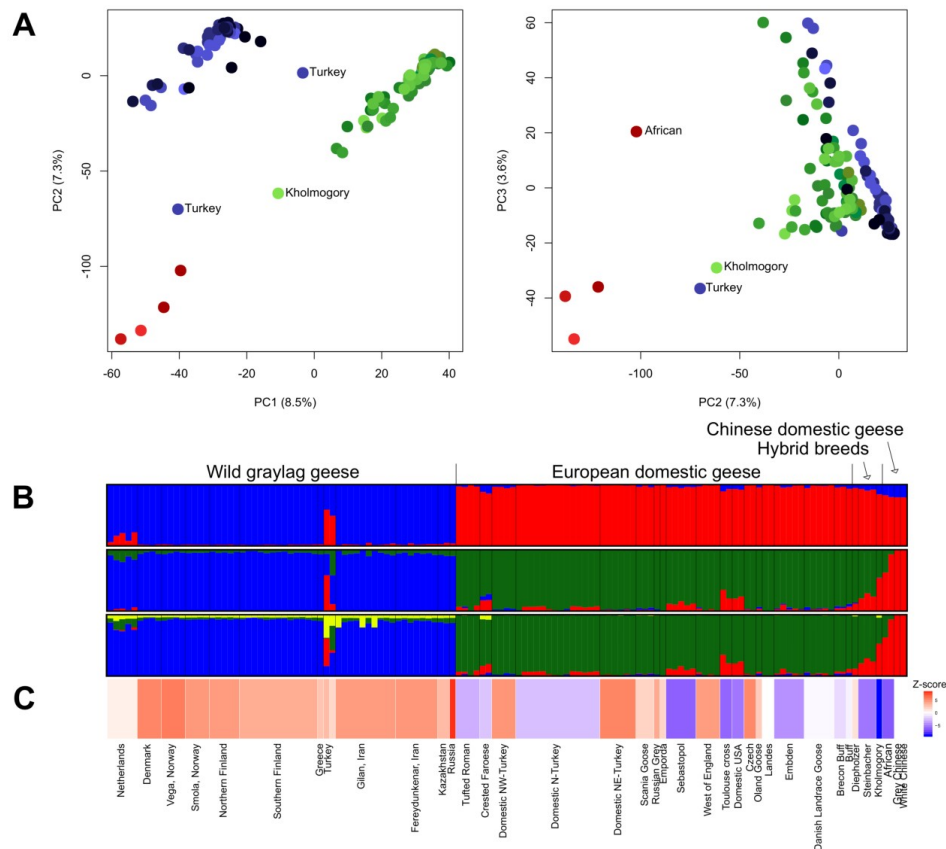


Figure 3. The genetic divergence and hybridization patterns in graylag and domestic geese. Population status and names labelled as in Supplementary File 1, Table S1. The colors in A) and B) are associated to different groups as follows: graylags (blue), European domestics (green) and Chinese domestics (red). A) The first three principal components summarizing the genetic variation in geese (percentage explained by each PC is shown). Different shades refer to different populations. B) STRUCTURE assignment plots for $K=2$, $K=3$, and $K=4$. Each vertical bar represents one individual with K number of colors indicating proportion of ancestry from the inferred clusters, and populations/breeds are separated by black vertical line. C) Plot relating to the f_3 (Supplementary File 1, Table S5) values obtained for each population. Turkey refers to two adjacent bars in the plot since the Turkish graylags were analyzed as two separate individuals. The more negative the f_3 , the more significant is Z-score in favor of admixture. The f_3 values were not calculated for Landes and the Chinese geese, as they were used as source populations, thus they were given an f_3 value of 0.